Text, Images and Language Translation in Facebook posts.

Andy Heath
Axelrod Access for All.
andyheath@axelafa.com

Rory Heap
Whitstability
whitstability@gmail.com

Abstract

This paper considers the accessibility of images and alternative text descriptions for them and language translation in posts made using Facebook. The paper looks at contexts where visual perception is unavailable and, considering Facebook as a system, makes recommendations for integrating simple tools supporting the provision of alternative text for images with particular application to text embedded in images.

Keywords:

Accessibility, Facebook, Images, Alternative Text, Screenreader, Optical Character Recognition, W3C/WAI, Web Content Accessibility Guidelines

Introduction

For many people Facebook is an important part of daily life, with many possible uses: it might be used to chat with friends, stay in touch with children or other relatives, participate in political debate, undertake political campaigning, find out what events are on in particular areas of interest, publicise professional events and achievements to colleagues and so on. Uses are legion. However, much of the content can be completely unusable by a proportion of the population by virtue of its visual nature or its visual encoding of information. The excluded population includes persons who are unable to perceive visual content by virtue of the context they are in or by virtue of a visual impairment, persons unable to understand text that is encoded in an image in a language in which they are not sufficiently skilled and various combinations and permutations of these. In this paper we look informally at some of the technical factors involved in posting and in understanding posts involving images and propose a simple mechanism whereby access to the content of posts might be extended to a greater number of people than currently. We assume the goal of access to the content of posts by the greatest

number of people and the consequent expansion of social contact to be a "raison d'etre" of Facebook and we assume that Facebook management would wish the platform to adopt practices that support that aim.

The problem

Facebook is a popular platform for social contact and a great many posts include images as part or all of their content. It is desirable that all people, including those who are not able, for whatever reason, to use visual perception to understand the content of images at that time in that context are able to perceive and understand the information content of all posts they are presented with. To achieve this it is necessary to present visual content in a non-visual way. A common way to do this is with the provision of textual alternatives for the visual content. Pure textual content can be easily rendered in auditory form (i.e. spoken) by the use of a screenreader, a technology (usually software) that supports turning textual and navigational information (for both interface and content) into auditory content and providing auditory cues when interfaces are used and content navigated. Sometimes tactile cues might be provided as well. There are a large number of different screenreader and related or similar technologies, some are embedded within the operating system and some are third-party provided, some operate with some devices and not others and with particular media technologies and software environments and not others and lack of interoperability across technologies is a frequent problem. The Apple iOS screenreader technology is called VoiceOver and the Android one is called Talkback. Third-party PC-based examples include JAWS For Windows and Window-eyes.

The common process is that when a screenreader technology process encounters an image or other visual object, if there is a provided textual alternative in a recognised form, it can turn that text into speech, effectively providing perception to a degree via the "equivalent" text. In fact providing text alternatives for non-text content (in this case visual) can support many other common modalities for consumption as explained in and recommended by the W3C/WAI Web Content Accessibility Guidelines 2.0 Guideline 1.1 "Guideline 1.1: Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, Braille, speech, symbols or simpler language."[1]

Images in Facebook posts are typically used in many ways and for many purposes. In almost all cases (a rare exception is discussed later in the paper) it is desirable that the information in the image is provided in an alternate form that can be perceived non-visually.

A common case is where an image contains embedded text: that is text that is bound into the image so that it can only be rendered in a

raster fashion, as part of the image not directly as text. The embedded text may be the entire information content of the post (and often is as we see below) or it may be part of the information and there may or may not be other directly textual information. Consideration of relationships between the information in the image and any other content in the post is a complex topic but it is safe at this point to say that usually the author intends any text in the image to be perceivable and understandable (or otherwise convey information, possibly by metaphor) to the consumer of the post. The difficulty with providing non-visual perception using the text in such a case is that the original text is often not directly available. Nevertheless, it is desirable that the text that is embedded be provided as part of a textual alternative for the image.

A technique that can sometimes be useful with embedded text is reconstruction of the original text from the image by means of Optical Character Recognition (OCR). OCR is not a perfect process; sometimes it succeeds, sometimes it partially succeeds (some text is reconstructed correctly but there are errors) and sometimes it fails, depending on the quality and features of the image, the text format, the actual text, the contrast, image and text colours and font and the processing applied. It is however, sufficiently good at recognising text to be used to recognise car registration plates on moving vehicles (admittedly a constrained vocabulary and format) in the UK Police Automatic Number Plate Recognition System [2]. There are also free online web services that will perform OCR on an image to extract any text that may be there, for example [3]. Some screenreader technologies can perform Optical Character Recognition on an image and some cannot. Relying on users having a screenreader that can do OCR to perceive the content of text embedded in images raises a number of problems:

1. We cannot rely on all users who need to do this having such a technology available for use in that particular context
2. There are significant interoperability problems in arranging that a technology can work in all contexts in which text with embedded images (and those images with possibly embedded text) are encountered
3. Some aspects of the relationship between the text and the rest of the content of a Facebook post can only be reasonably addressed at authoring time not consumption time (we elaborate on this later in the paper)
4. Not **every** user who can benefit from the proposal we make in this paper for provision of support for alternative text for images will be using any screenreader technology at all

In this paper we will argue that integration of OCR support in a simple interface invoked at image publishing time combined with support for authoring of displayable Alternative Text for images that describes the content of the images would significantly enhance the

accessibility and the audience-reach for many Facebook posts. Whilst the main driver for this paper is the problem of embedded text the tools, interfaces and services for supporting authors to create alternative text for non-textual image content are very similar to those for dealing with embedded text and so we deal with both of these things in an integrated fashion. A main contextual driver is the situation where visual perception is limited or not available (as for visually impaired people or situations where vision cannot be used) but the benefits of what we propose extend to consumers of **all** posts containing images whether visual perception is available or not.

Our main argument is built around two use-cases: the general case where visual perception is not available and the case where visual perception may or may not be available but translation from one language to another might be necessary to understand a post but there are several technical use-cases that relate to this and we consider what are the implications of each.

In passing we point out that the value of constructing alternative textual descriptions of images does not only extend to contexts where visual perception is not available or where language translation is needed. The notion of displaying text instead of an image can be extremely useful where display space is limited or where download bandwidth is limited, where one might choose not to display images at all. We do not discuss this particular context in detail here but just note its relevance.

The Evidence and Argument

A very informal study of the most recent 100 posts shown in one of the author's own Facebook News feeds yielded the following results:

- Number of posts consisting entirely or mainly of an image with embedded English language text: 20

- Number of posts consisting entirely or mainly of an image with embedded Spanish language text: 6

- Number of posts consisting entirely or mainly of an image without embedded text: 45

- Number of posts consisting of just English language text in pure textual form (i.e. containing no images):   16

- Number of posts consisting of just Spanish language text in pure textual form (i.e. containing no images):   3

- Number of posts consisting of just Swedish language text in pure textual form (i.e. containing no images):   1

- Additionally, of those posts with embedded text in an image (either in English or in Spanish), in 20 of them the embedded text was either the entire meaningful content of the post or the major part of that.

- 71% of the posts contained an image – far more posts than pure textual ones.

- None of the posts with text embedded in an image had explanatory text reproducing the embedded text.

- None of the posts containing an image without embedded text had any content describing what was in the image that was identified as such

- The purpose of the image differed in different posts and included (but was not limited to) the following factors:

  - In some cases the image was directly illustrative in relation to textual content also in the post, and not essential to understanding the post though there was no indication of that fact in the content so a user without visual perception would not know it.
  - In some cases the image was allegorically or metaphorically illustrative in relation to textual content in the post (such as a humorous cartoon illustrating serious textual content). Whilst such a post might be understandable without knowing the image content it is a poorer experience and the more so because a user without vision could not be aware that the image wasn't essential to understanding the post.
  - In some of the posts with images without embedded text knowledge of the image content was essential to gaining any understanding the post – for example a post with the text "Look who the wind blew in" and an image of a person with no other indication in the text of who the person was.
  - In many posts the situation was between the two extremes, the meaning of the post being conveyed partly in the text and partly in the image but without any indication of that.

Note that we make no claim that this data is unbiased or shows that the problem is pervasive but give it here for its provision of examples for discussion. Nevertheless, we believe the problem **is** pervasive and that a rigorous study would show that.

This position leads to inaccessibility for many individuals. For example, in the case of people without any visual perception many

posts containing an image without any description of the image would be completely meaningless, and in those cases where the image is directly illustrative or its meaning fully or partially reproduced in the accompanying text but that fact not made explicit the quality of the experience would be substantially reduced from what it might otherwise be. The overall effect with the posts in this informal survey is that only 29% of the posts were fully accessible to people with visual impairment sufficient to prevent perception of the image content. Neither do these figures take account of images posted in comments on posts.

As we have said, a great many posts consist entirely or mostly of text embedded in an image. There are different reasons for this. Our purpose here is not to categorise them but it is worth looking at some real examples to better understand the problem. One case shows in Figure 1.

Figure 1. Example of an image that was posted on Facebook consisting of a photo with embedded text. The image shows a photo of a sign in a Doctor's surgery. The sign has the text "OPEN SURGERY, ONE PROBLEM ONLY, This is for the quality and safety of patients. We also want to avoid any inconvenience and delays to other patients waiting."

In this case the image that was posted is a photograph (taken by an author of the paper) of a sign containing text and so the image naturally has the text embedded. In its current form, without accompanying pure text, the image is not accessible to a sufficiently visually impaired person (or in a context where vision cannot be used). In this case, since the original source is itself a photo, its unlikely that a user will provide separate text unless supported in doing so by integrated tools since the original text wasn't available to the photographer and doing so would involve separately typing the text.

It should be noted that not only is this embedded text not accessible to anyone without visual perception it isn't available either to any person without sufficient English language skills to read and understand it as it is. Additionally, because the text is locked into the image, language translation tools are not available to be used on the text. Its perception is limited to English-speaking people with sufficient visual perception. This is an unnecessary limitation and there are easy ways to circumvent it but they require extraction of the text from the image.

6

Posts of images like the one in Figure 2, essentially a screen-grab, seem to be increasing in prevalence as the tools for easily creating them and copying them around become ubiquitous. In such a case it might not be

Figure 2: The image is a post showing an iPhone text exchange probably gathered by a screen capture on the device. The user of the phone is called "Mum". The text of the exchange is "Other: Finally, you've entered the digital age and got a smartphone!; Other: How is it?; Other: Mum?; Other: Helloooooo?; Other: Why aren't you answering??; Mum: Howdoyoudoaspace"

impossible for its originator to copy the original text and post that instead or as well but it isn't as easy for the user to do that and doing so wouldn't have the same visual impact – the picture makes it immediately visually clear that the conversation is taking place in the texting environment on an iPhone but from the text alone this would not be apparent. The variety of environments like this, which involve text, is so great that establishing any immediate automatic interoperability across them directly using the text would be very difficult (though many systems go some way towards this by providing cut/copy/paste and sharing functionality). Posting an image also sidesteps in practical terms for the user any issues of Intellectual Property and "who owns the content". For both of these reasons images like this one containing text are likely to continue to increase in ubiquity.

Another case of embedded text occurs where authors add text to an image, as in Figure 3. This kind of post is extremely common probably because tools for manipulating images and adding text annotations have become easy to use and ubiquitous. The author of the picture clearly had

Figure 3: Picture of Darren Hall, UK Green Party election candidate for Bristol West in what is probably a modern hotel lobby, overlaid by the text "We will prove there is a party that stands up for the people long-abandoned by the politicians of the establishment. Darren Hall, Bristol West"

the text available but may well have been unaware of the need to ALSO make that text available for users with a context in which visual perception is unavailable and as potential input to translation tools. Regrettably this applies to a very large majority of Facebook postings and it is likely to continue to be the case that authors do not readily think about the non-visual contexts.

Figure 4: Shows an image of a Facebook post containing the text "Arduo, diria yo…" as pure text followed by a link "Show Translation", followed by an image containing the spanish text "LAS OPORTUNIDADES NO SON PRODUCTO DE LA CASUALIDAD, MAS BIEN SON RESULTADO DEL TRABAJO. WWW.REGAECR.COM"

Finally, lest it isn't already clear from Figure 1, embedded text is not easily

7

available for input to translation tools, meaning that an image with embedded text is only understandable to persons who can both perceive the image visually and understand the language in which the text is written. Figure 4. Shows an image of a Facebook post containing the Spanish text "Arduo, diria yo…" as pure text and an image with some embedded text which carries the main meaning of the post. Within the Desktop browser-based version of Facebook pure text can usually be translated to the reader's native language. For a sighted user this functionality is invoked simply by clicking on a link labelled "See Translation" and the translated text is placed inline in the post. Regrettably the substantial text inside the image is not available for translation, so the content of the post is not accessible to many native English speakers, including, in this example, an author of this paper. The experience with at least one screenreader (tested with Jaws 8.1 on Windows XP using Firefox 26.0, 26th February 2015) is similar: translated text is placed inline but the text in the image remains inaccessible.

Whilst the author of the post might have also posted the original text, in practice this may not happen often. It is open to a consumer or sharer of the post to save the image outside of Facebook and upload it to an OCR service to extract the embedded text then use that but in practice:

1) This is tedious to do and probably will not happen often

2) A consumer without visual perception might not know that there is embedded text anyway so would not know to do that and

3) A consumer without visual perception is not in a position to know whether the text extracted by OCR is correct: the best person to know this is the original author, or in some cases a person **with** visual perception who is sharing the post (who can know if the extracted text is correct but not know the author's intentions) and

4) The original author is in the best position to determine whether extracted text should be included in textual alternatives and what, if any, other text needs to be included to understand the intended information content of the image.

What is needed then are tools to help the original author, and perhaps a later sharer, deal with these issues.

Towards a solution

Technical Context

In this paper we address the Facebook Desktop web-based software i.e. as a user might interact with Facebook using a browser such as

Firefox on a laptop or desktop computer. We do not address mobile devices such as phones or tablets. In fact at the time this paper was originally written the authors found it disappointing that the translation functionality was not implemented in mobile platforms such as iOS (as tested in iOS 8.1.3 26th February 2015). It has since appeared in those platforms, which is pleasing.

We make no attempt at all here to discuss the accessibility of a posting interface, which itself is an interesting topic worthy of further serious consideration.

Standards Context and Use Cases for a System

There are some useful sources in standards and standards-related work that provide some guidance on the use of Alternative Text for images. Whilst the authors don't recommend users are directly referred to them or required to use them in any way in deciding what to put in Alternative Text for an image, because such may be counter-productive and discourage authors from providing alternative text at all, we **do** recommend that implementers read them for guidance in understanding what is required. We also lean on them to affirm our views on what use cases are important here. Two sets of documentation of note are from W3C [4] and Wikipedia [5]. Both of these are useful for a technical person to gain an understanding of how alternative text can be used. There is also an ISO/IEC Technical Specification [6]. This last is complete and in the opinion of the authors contains very useful advice but unfortunately it is not free. An implementer might wish to summarise some of the advice in these sources and include such as "help" material for Facebook.

Use Cases and their Handling

The following use cases appear to be particularly relevant to Facebook image postings:

1. A user posts an image having information content that is not otherwise in the post
2. A user posts an image where the information content is also described in the text of the post either identically or as text with the same meaning – i.e. the image is completely illustrative to the post
3. A user posts an image that is entirely decorative and contains no information content relevant to understanding the post. Note that at time of paper writing the authors could find no example of such a case in Facebook postings and it could be that it is a rare usage
4. A user posts an image where displaying alternative text would interfere with the purpose of posting the image or making the post

We propose handling each of these in different ways.

- For the first (1 above), it is clear that the addition of alternative text describing what is in the image will enhance the accessibility of the post.

- In the second case (2 above) knowing the image content is not essential to understand the post. However, knowing an image's content may substantially improve the experience of the consumer. The best solution for the consumer would be for alternative text describing the image content to be provided even though that information is also in the text. On the other hand, if the author were not willing to provide image description then it would be better to have an indication that the image content is not essential to understanding the post than for the consumer to have no information about the image at all. We recommend the author or sharer be encouraged in the structure of the interface to **always** provide alternative text describing image content and that the possibility to provide alternative text that says just "Image is illustrative only" or some similar phrase is mentioned in hyperlinked guidance material.

- The third case (3 above) may be extremely rare. We recommend the same practice is followed as recommended for the second case: that the author be encouraged to always provide alternative text description but the possibility that the description be "Image is purely decorative" be mentioned in hyperlinked guidance material.

- The fourth case (4 above) is more complex to deal with. One way in which this case can occur is with visual jokes. Consider figure 5.
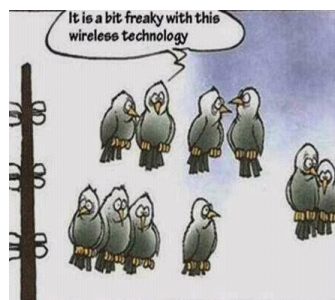

Figure 5: An image of birds apparently sitting on a telegraph wire next to a telegraph pole but no wire is present. One bird is saying to another "It is a bit freaky with this wireless technology"

In figure 5 a visual joke is made. If alternative text is visible to a consumer of the post then it might "spoil" the "visual" joke. There are many ways such "spoiling" can happen, not necessarily where text is involved. The ideal for this case is where a user has requested alternative text we would be able to display the embedded text as pure text **and** a description of the image but for a context where alternative text has not been requested the alternative

and embedded text is hidden. We propose below a mechanism that can do this.

Two additional use cases involve posting of URL's where Facebook searches the referenced content for an image to render and posting of images with external sharing tools such as photo tools on iOS or Android. These are worth of attention but we do not address them here in order to save space and not detract from our main argument.

It should be noted that whereas it isn't difficult to deal with questions such as "should alternative text be displayed to a user in this case" or "what should the alternative text be" at post authoring[1] time, it can be difficult to do so at consumption time because doing so often requires knowledge that only the original author has. The process we will propose to deal with these requirements should therefore run at authoring time (when someone posts an image). However, there is value also in running the process when someone shares an existing post containing one or more images. Its value here is in supporting third parties in making repairs (supplying accessibility information for images where it was not previously supplied).

Implementation

The prime consideration must be to preserve usability. The provision of accessibility information must add only a few extra clicks to the posting process and must require the least additional work as possible from a producer or sharer. If this is not achieved then the process is unlikely to be used. Ideally such a mechanism will make its presence known unobtrusively to a user making an image post, in a way that the user is not slowed unless the user chooses to supply the information. We have designed the following process with that requirement strongly in mind.

The first stage is when a user wants to update his or her status and wants to include one or more images in that update. The user will see something like Figure 6.

---

[1] Post-authoring as in the time when a post is made, NOT as in "after authoring"

Figure 6: Figure 6: Modified Facebook posting interface showing hyperlinks to Accessibility Information below each of two uploaded images. Each image contains embedded text but the actual text content is not meaningful in the context of this paper, only the fact that the text is present is relevant for our purposes here

As the user uploads an image a hyperlink appears beneath it leading to a pop-up screen to enter that information. If the user decides not to enter any accessibility information posting proceeds as normal and the user is not troubled further. The pop-up screen for entering accessibility information for each image is similar to the following (Figure 7):
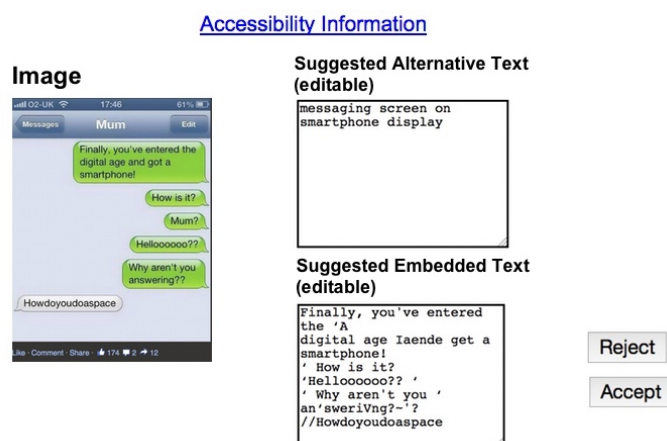

Figure 7 Mock-up of a possible interface screen to enter accessibility aspects of an image. Includes the submitted image which is a picture of a smart phone messaging conversation and two editable text widgets, one labelled "Suggested Alternative Text (editable)" containing the text "messaging screen on smartphone display" and the other labelled "Suggested Embedded Text (editable)" containing the text "Finally, you've entered the 'A digital age laende get a smartphone! ' How is it? 'Helloooooo?? ' ' Why aren't you ' an'sweriVng ~'? //Howdoyoudoaspace". There are two action buttons, one labelled "Reject" and the other labelled "Accept"

In the example shown in Figure 7, two automated detection services have been run automatically. For this paper each service has been run manually and the output captured. Each output is produced in an

editable content input box so that the user can change the suggested text.

The Suggested Alternative Text has been produced automatically by the alt-text-bot service [7] which is a service run on Open Source software that describes the content of images posted to twitter. Services like this cannot be perfect because even if they detect image features correctly they are not inside the mind of the originator. However, it has done a good job in this case and the alternative text it suggests is acceptable.

The Suggested Embedded Text as stated has been produced by the Free OCR service referenced here [3]. The text produced which shows in the Figure description above is not perfect but editing it to make it perfect is an easier job than typing it all, and in fact it may be "just good enough" to be understood without any editing.

We have not in this example provided separate Accept or Reject controls for each of the suggested texts though that could easily be done. Instead the idea here is to recommend the practice of clearing the editable input box to reject that text without adopting anything in its place.

The same interface needs to be called up when someone shares an existing post containing an image but with different conditions. In the case of sharing a post with an image in the automatic generation should be invoked **only** if the respective field is empty but any known values for the two fields should be displayed to the sharer with the opportunity to change the values of text provided with the image as appropriate. If no text is already supplied in a field then the automatic generation should run to provide an initial guess. If there is text in the field already it does not need to be re-generated and in fact if there is text which has been edited since automatic generation it might be a strong indicator that the original (or previous) poster or sharer has supplied text or edited auto-generated text. It would be a useful practice then in an implementation to record whether the value of each of these fields has been changed since auto-generation.

Render or consumption time

At consumption time it is desirable that alternative text and/or embedded text is only displayed if requested. This addresses the situation where the purpose of posting an image would be damaged by revealing alternative text for the image. A common case of that, in a humorous cartoon, is shown in the examples above.

There would appear to be two ways that display only on request can be achieved. One such way is to provide an individual preference setting for displaying alternative text. The displaying software could

consult that setting at consumption time for the post and the content of those fields displayed or not as appropriate. A benefit of this approach is that a user not requiring display of alternative or embedded text can see exactly the same interface as at present.

Another way to achieve the same effect, which is quite simple, is to provide a hyperlink accompanying an image similar to the "See Translation" links. This might be entitled "Accessibility Information" and selecting it at consumption time might place inline in the post any alternative text and any embedded text that is known for an image. Such a link is probably best placed **before** any "See Translation" links because where a consumer wishes content to be translated and has requested display of alternate and embedded text the consumer is very likely to want the content of the alternative and embedded text translated as well. By placing the "See Translation" link **after** the "Accessibility Information" link the translation process can easily operate on the alternative and embedded text content as well as the direct text content in the post. Where there are several images in a post some careful thought needs to be given to choosing a layout and in such a case it would seem sensible for Accessibility Information links to immediately follow each image but for a "See Translation" link to be at the end of a post and apply to all of the post rather than a segment of text. Similar arguments apply to images in comments.

Conclusions and Future Directions

The major theme of this paper has been to suggest a mechanism which might be used by some authors with very little additional authoring-time load to make images, particularly those containing embedded text accessible to a wider range of people and in a wider range of contexts than they currently are. The mechanisms suggested to do this have been by enabling an automatic image content or scene analysis service and an automatic Optical Character Recognition service to be used at authoring time and at share or repost time. We have suggested a simple interface design to support integrating such services in what we believe to be a fairly useable manner.

This extension of access has included exposure of embedded text to language translation tools, thus increasing understanding of posts across different languages. But these are not the only ways in which such a mechanism might extend access. By providing a lightweight mechanism by which attention is drawn to the need to provide alternative text as description of the content of images we hope that authors might be encouraged to provide such description and increase awareness of accessibility needs for all authors.

Optical Character Recognition services and scene or image feature analysis services are yet a long way from perfect but they are

improving and some tools are already able to provide substantial support in describing images and embedded textual content within images in a "just good enough" or "nearly good enough" fashion. Their use is likely to assist in making Facebook posts accessible to a wider range of people and we would like to see Facebook take up their use sooner rather than later in order to reap the benefits of their use.

Acknowledgements

Martín Szyszlican, Sinapsis UX Research Team (http://accesibleyusable.com/sinapsis/team/), martinsz@gmail.com for critical review and very helpful comments.

References

[1] http://www.w3.org/TR/UNDERSTANDING-WCAG20/text-equiv.html visited on 6th May 2015

[2] http://www.police.uk/information-and-advice/automatic-number-plate-recognition/ visited on 26th February 2015.

[3] http://www.free-ocr.com/ visited on 26th February 2015.

[4] HTML5: Techniques for providing useful text alternatives http://www.w3.org/TR/html-alt-techniques/ visited on 26th February 2015, this is a public draft and should not be considered final. A later editors draft is available but this should be considered even less stable and therefore probably not used.

[5] http://en.m.wikipedia.org/wiki/Wikipedia:Alternative_text_for_images, visited on 26th February 2015

[6] ISO/IEC SPECIFICATION TS 20071-11 Information technology — User interface component accessibility — Part 11: Guidance for alternative text for images. Available from http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59423, visited on 26th February 2015. Not free.

[7] The alt-text-bot service is hosted at http://alttextbot.com/ visited on 29th June 2015